DESIGN OF A SONAR SYSTEM FOR VISUALLY IMPAIRED HUMANS

Cameron Morland

Boston University
Department of Cognitive and Neural Systems
677 Beacon Street, Boston, MA 02215, USA
cjmorlan@bu.edu

David Mountain

Boston University
Department of Biomedical Engineering
44 Cummington Street, Boston, MA 02215, USA
dcm@bu.edu

ABSTRACT

Constraints and features useful for an effective and easy to learn human sonar device are described. These include matching spatial cues generated by the device to those in the world. Techniques used by natural echolocators, including specifications of signal type, emitter, and receiver are briefly reviewed, as is techniques of converting ultrasonic signals to the audible range and techniques for externalizing sounds. Finally, a prototype sonar system designed while considering these ideas is described.

1. INTRODUCTION

Echolocation is a method of perceiving the world. Echolocating animals emit noises then listen to the reflections or echos of these noises. It is used for hunting by insectivorous bats and dolphins. Fruit bats [1], certain cave-dwelling birds [2], and small mammals [3, 4, 5] echolocate for purposes of navigation (collision avoidance, orienting).

Some visually impaired humans have long used echolocation as part of their orienting repertoire, sometimes unconsciously [6]. They generate noises by tapping a cane or foot on the ground, snapping fingers, or clicking the tongue. A very small number of humans are skilled echolocators who echolocate intentionally. There has been some effort to train people to echolocate [7].

This paper introduces a device which makes echolocation more effective and easier to learn, while being audible only to the user. When completed the device will cause nearby objects to appear to be emitting sounds.

Section 2 discusses a number of factors which were considered important features for a human sonar device. Section 3 considers how some of these factors have been met by natural echolocators, and important issues for implementing these synthetically. Finally, section 4 describes the workings of a particular human sonar device prototype.

2. DISPLAY DESIGN CONSTRAINTS

A number of properties are desirable or necessary in order to make a sonar system which is useful but has a sufficiently shallow learning curve.

2.1. Minimal Interference with Other Senses

Visually impaired people are known to rely more strongly on their remaining senses. To prevent the use of this device from becoming a trade-off of benefits and inconveniences, this device should only minimally block or otherwise diminish the effectiveness of the user's senses.

2.2. Externalization

Sounds presented over headphones tend to sound as if they are inside the head [8], or coming from the surface of the head. Since the goal of this project is to make the sounds appear to be coming from the perceived objects, as if they were themselves emitting the sounds, a method for externalization (such as a head related transfer function or HRTF) should be applied to the synthesized echoes.

2.3. Spatial Cues

To reduce the degree to which users must learn to use the device, the spatial cues provided by the device should match, to as large a degree as possible, the cues that are available under natural hearing. That is, if an object is both emitting sounds of its own and reflecting sounds generated by the device, the differences in interaural timing and interaural level should be similar to allow discrimination of azimuthal angle, and similar spectral filtering should be present for discrimination of elevation.

3. BACKGROUND

3.1. Emissions

Three major classes of emissions are used by echolocating animals: clicks, frequency-modulated sweeps, and constant frequency tones. These three emissions, cartoons of which are shown in Figure 1, are used by species which occupy different ecological niches and have different perceptual requirements.

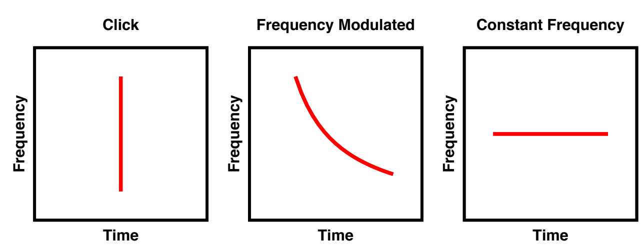

Figure 1: *The three types of echolocation sounds used by animals.*

3.1.1. Clicks

An ideal click is the Dirac delta function, having zero duration and equal power at all frequencies, but real clicks inevitably have finite bandwidth and non-zero duration. Animals which echolocate in air generate the click by a tongue-click, and that is a method recommended for visually impaired humans endeavoring to learn to echolocate [7]. Animal-generated click signals usually have a duration of less than two periods of the dominant frequency (for example see emissions of *Steatornis caripensis* [2] or *Rousettus aegyptiacus* [9]). Natural clicks of *Rousettus* have been fitted by Gabor functions [9].

The short duration of clicks makes them ideal for precise timing, but limiting the duration requires a higher peak power to put a given amount of energy into the emission. So their range is limited.

3.1.2. Frequency Modulated Chirps

Frequency Modulated (FM) chirps could include a wide variety of signals, but the ones actually used by animals are fairly specific. They are used exclusively by microchiropteran bats using the larynx. Most, if not all, FM chirps are downward sweeping harmonic complexes [10]. Further, they are very closely approximated by signals having linear period modulation (hyperbolic frequency modulation). These signals are ideal for cancelling Doppler distortion effects caused by high target velocity [11, 12] which allows accurate timing information while using a long duration emission. For a fixed peak power level, a longer signal allows more energy to be transmitted in each emission, which increases the usable range or effectiveness in the presence of noise.

3.1.3. Constant Frequency Tones

An idealized tone exists for all time, and has zero bandwidth. Since the signal does not change over time, it provides no timing information. In practice therefore constant frequency (CF) emissions always include a click or rapid frequency sweep at one or both ends of the emission. Certain bats echolocate using CF emissions to hunt insects, emitting tones up to several tens of milliseconds in length. They have evolved a cochlea which is “foveal,” meaning that a large portion, sometimes as much as half, of the cochlea is devoted to a narrow frequency band centered around the frequency of the emissions [13]. This specialized detector provides sufficient sensitivity to allow the bats to detect Doppler shifts, including those causes by the wing-beats of prey insects.

3.2. Emitter

As mentioned above, microchiropteran bats generate emissions using their larynx, while others use a tongue click. These then propagate through the mouth and nose, and can give complex beam patterns. The -6 dB beam widths can be quite wide, 120° or more, as in *Hipposideros terasensis* [14] (a CF bat), or fairly narrow, less than 60°, as in *Eptesicus fuscus* [15] (an FM bat).

3.3. Receiver

Two ears allows for the generating of interaural time and level differences, which are used as azimuthal cues. Furthermore, the external ear is shaped in a complex manner, which causes the incoming signal to be filtered differently, depending on the elevation of the sound source (the head related transfer function, HRTF).

In the big brown bat, *Eptesicus fuscus*, which echolocates using FM chirps, the HRTF spectrum is remarkably systematic. This species has just one major notch within its range of hearing, and this notch moves smoothly with elevation; it is unchanging above midline, and moves monotonically down in frequency below the midline [16].

Directionality of sensitivity can also be described, and affects where the animal can detect obstacles. The ears of *Eptesicus fuscus* point somewhat out to the side, and are particularly sensitive directly along their axis. When a plot of the power of the emission is combined with the receiver sensitivity the result is the strength of the signal reflected back to the ear by a target. *Eptesicus fuscus* has been shown to have a cone of $\pm 30^\circ$ [15].

3.4. Signal Processing

In the synthesized echo-locating system an ultrasound signal is played by the emitter, reflected by an object, and received by the microphones. For this artificial sonar system, the primary signal processing task is to shift the frequency of the received echoes into the audible domain for presentation to the human user. There are two common ways of doing this, heterodyning, and time stretching.

The effects of these two techniques are contrasted in Figure 2.

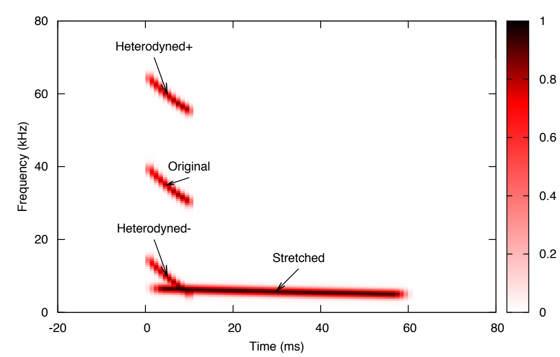

Figure 2: *Effect of frequency shifting by heterodyning and time-stretching on a typical frequency modulated sonar chirp.*

3.4.1. Time Stretching

Time stretching involves recording a signal of a certain length, then replaying it at a slower speed. Time stretching has been used in a number of human sonar devices, including instrumented human divers [17, 18].

Given a sinusoidal input signal $\sin(f_s 2\pi t)$, replaying it slowed down by a rate r, gives

$$s = \sin\left(\frac{f_s}{r} 2\pi t\right) \quad (1)$$

Thus slowing down by r causes the output signal to have its frequencies scaled down by r.

As is clear from Figure 2, time stretching reduces the bandwidth (linear) of a signal while maintaining the number of octaves (logarithmic) spanned by the signal, as well as spectral relationships within the signal. Since the auditory system has a log-frequency response, a time-stretched signal will sound the same as the unstretched signal, except for timing; it is equivalent to transposition in music. The timing of the signal, of course, will be affected.

3.4.2. Heterodyning

Heterodyning linearly shifts the frequency of a signal, using the identity

$$\sin\theta\sin\phi = \frac{\cos(\theta-\phi)-\cos(\theta+\phi)}{2} \tag{2}$$

Let $\phi = f_h 2\pi t$ and $\theta = f_s 2\pi t$, where f_h is the heterodyning frequency and f_s is a component of the signal. This splits the incoming signal $\sin(f_s 2\pi t)$ into a high frequency component shifted up by the heterodyning frequency

$$s_+ = -\cos((f_s + f_h)2\pi t) \tag{3}$$

and a low frequency component shifted down by the heterodyning frequency,

$$s_- = \cos((f_s - f_h)2\pi t) \tag{4}$$

Figure 2 shows that the bandwidth of the signal is unchanged, while the number of octaves spanned can increase considerably (in the case of shifting a signal to 0 Hz, the number of octaves spanned becomes ∞). The log-frequency nature of the auditory system will therefore cause a heterodyned signal to sound stretched in frequency. For example, a signal with 1 kHz bandwidth between 41 and 42 kHz (just 0.035 octaves) heterodyned down by 40 kHz, will cover 1-2 kHz, which is one octave. Frequency fluctuations, even small ones such as those caused by Doppler shifting, can be detected much more easily due to this octave stretching.

3.5. Earphone Design

Earphone design is primarily affected by the need to externalize the signal, but the constraint of minimizing impact on natural hearing also had an impact.

Multi-channel headphones (e.g. those by S-Logic, [19]) have been created which present different sounds from different angles around the ear. The sound is thus filtered by the individual's pinna transfer function to give an impression of vertical dimension and of being external.

There are several problems with these headphones for the purposes of this project. Bulkiness and weight are the most obvious problems, since the device should be wearable for an extended period of time without fatigue. More importantly, these headphones block natural hearing. Further, they introduce their own reverberations within the headphone (acoustic felt notwithstanding), which could mask some closely-spaced reverberations from the environment, reducing the acuity of the instrumented human.

Ordinary closed insert earphones prevent the externalization of sounds, apparently due to sound waves leaving the ear canal being reflected back in by the plug. Special earphones which leave the ear canal open, combined with even very crude HRTF-like filtering, can allow an impression of externalization [20]. In particular, reassembling just the first 16 components of the Fourier transform of the spectrum of the HRTF gives a smoothed but sufficient HRTF.

4. METHODS

The above sections give a number of constraints on how a human sonar system should work, and context of how these constraints can be met. This section describes how this prototype device has been constructed. Since experimentation is only now beginning, many parts of the design have not be validated and are subject to change.

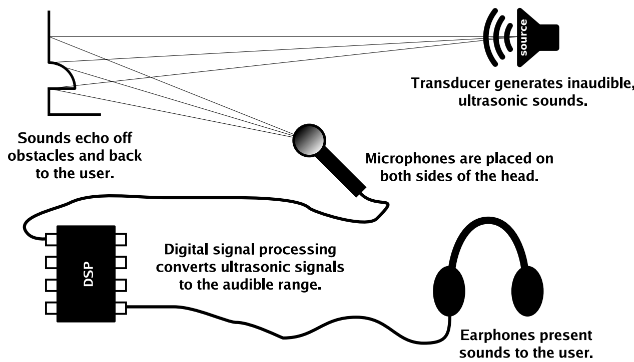

Figure 3: *Outline of operating of the device.*

A schematic of the device is shown in Figure 3. Ultrasonic sounds around 40 kHz are emitted by a piezoelectric transducer. These reflect off objects in the world and return to the user, where they are picked up by miniature microphones mounted near the ears. Digital processing converts the received signal into the audible range, which is then presented using special earphones. The device works by very specific, but limited, processing of the received echo signal to retain the spatial cues.

4.1. Emissions

Research is ongoing into emission choice. Informally, clicks have been shown to give a sense of azimuthal location, while tones provide information about texture due to frequency fluctuations resulting from Doppler shifts. In general the signals have been centered close to 40 kHz to match the maximum power output of the emitter.

4.2. Emitter

The emitter used in the prototype is a SensComp Piezo Transducer 40LT16 [21], a transducer designed for transmitting ultrasonic signals in air. This transducer has a -6 dB bandwidth of 2.0 kHz, with its max at 40 kHz, and a total -6 dB beam angle of 55°. This beam angle puts the device in line with the beam angle of *Eptesicus fuscus*, discussed in Section 3.2.

4.3. Receiver

Two Knowles FG-3329 microphones are mounted on the head, one above each ear. These are connected to custom pre-amplifier boards which high-pass filter the signals above 14 kHz. The outputs of the amplifier are connected to a National Instruments data acquisition card. These microphones have a flat frequency response up to 10 kHz, and although they gradually become less sensitive above this point, they are still sufficiently sensitive around the 40 kHz range of interest.

Their directionality at such high frequencies has not yet been measured. However, since their diameter (2.59 mm) is smaller than the wavelength of sound at 40 kHz (8.5 mm), they will be fairly non-directional. The directionality of the transmitter and the effects of head shadow, therefore, will dominate the directionality of the system.

Placement of the microphones near the ears allows the interaural time differences to be identical to that of natural sound sources. The head shadow gives interaural level differences similar to those found naturally. One difference is that the entire range of frequencies within the head shadow are affected, not only the "high" frequencies, since the acoustic signal is entirely high frequency.

4.4. Signal Processing

Processing is done digitally on a desktop PC running GNU/Linux. Signals are recorded at 192 kHz at 12 bit resolution using a National Instruments PCI-6110 data acquisition card. Playback occurs at 48 kHz at 16 bit resolution through a sound card. All processing is performed on 64 bit floating-point numbers.

Timing information (e.g. for interaural time differences) is considered highly important for this device, whereas maintaining harmonic relationships is much less so. Heterodyning was therefore determined to be the preferred technique for frequency shifting.

Incoming signals are heterodyned by a 38 kHz sine wave. They are then digitally filtered. This digital filter introduces a generic (non-individualized) HRTF-like shape, as needed for externalization, and flattens the frequency response of the earphones. The signal is then low-pass filtered to remove aliasing before the 48 kHz output stage.

In bats enough of the transmitted signal reaches the ear to require the animal to reduce its hearing sensitivity [22] to prevent masking or hearing damage. This high level of transmission is unnecessary, but some signal reaching the ears allows the auditory system to measure the timing between transmission and echo reception. Due to the mechanical separation of the components in the synthetic echolocation system, the directionality of the emitter, and the head shadow, hardly any sound moves directly from emitter to receiver. It may be necessary to introduce information about outgoing signal timing to the earphone signal.

4.5. Earphones

This prototype uses Etymotic ER-3 earphones connected to open-ear inserts designed for surveillance. Since these earphones are designed to be connected to a sealed ear canal, they suffer from considerable attenuation of the low frequencies with an open ear canal. This attenuation is cancelled in the filtering stage.

4.6. Miniaturization

Most of the device, including the transmitter, microphones, amplifiers, and battery pack, are quite portable, and are in fact mounted on a baseball cap. Currently the processor, a desktop PC, is rather large and cumbersome. The intention is to replace this computer with a Blackfin Digital Signal Processor, which is small, low power, and can easily be configured with 24 bit digital to analogue and analogue to digital converters operating at 192 kHz.

5. CONCLUSIONS

A device designed to allow a human operator to perceive the location, texture, and motion of objects in the world through echolocation is described. Informal testing shows humans are sensitive to azimuthal location and velocity, depending on the type of emission used. Testing of the device, parameter optimization, and design is ongoing.

6. REFERENCES

[1] D. R. Griffin, A. Novick, and M. Kornfield, "The sensitivity of echolocation in the fruit bat *Rousettus*," *The Biological Bulletin*, vol. 115, pp. 107–113, August 1958.

[2] Donald R. Griffin, *Listening in the Dark: the acoustic orientation of bats and men*, Yale University Press, 1958.

[3] K. A. Forsman and M. G. Malmquist, "Evidence for echolocation in the common shrew, *Sorex araneus*," *Journal of Zoology, London*, vol. 216, pp. 655–662, 1988.

[4] Thomas E. Tomasi, "Echolocation by the short-tailed shrew *Blarina brevicauda*," *Journal of Mammalogy*, vol. 60, no. 4, pp. 751–759, November 1979.

[5] Edwin Gould, "Evidence for echolocation in the tenrecidae of Madagascar," *Proceedings of the American Philosophical Society*, vol. 109, no. 6, pp. 352–360, December 1965.

[6] Michael Supa, Milton Cotzin, and Karl M. Dallenbach, ""Facial vision": The perception of obstacles by the blind," *The American Journal of Psychology*, vol. 57, no. 2, pp. 133–183, 1944.

[7] Daniel Kish and Hannah Bleier, "Echolocation: What it is, and how it can be taught and learned," `http://www.worldaccessfortheblind.org/docs/echolocationtrainingguide.pdf`, 2000.

[8] WM Hartmann and A Wittenberg, "On the externalization of sound images," *J Acoust Soc Am*, vol. 99, pp. 3678–88, 1996.

[9] RA Holland, DA Waters, and JM Rayner, "Echolocation signal structure in the Megachiropteran bat *Rousettus aegyptiacus* Geoffroy 1810," *J Exp Biol*, vol. 207, pp. 4361–9, 2004.

[10] Martin K. Obrist, Ruedi Boesch, and Peter F. Flückiger, "Variability in echolocation call design of 26 Swiss bat species: consequences, limits and options for automated field identification with a synergistic pattern recognition approach," *Mammalia*, vol. 68, no. 4, pp. 307–322, 2004.

[11] Jan J. Kroszczyński, "Pulse compression by means of linear-period modulation," *Proceedings of the IEEE*, vol. 57, no. 7, pp. 1260–1266, 1969.

[12] J. Yang and T. K. Sarkar, "Doppler-invariant property of hyperbolic frequency modulated waveforms," *Microwave and Optical Technology Letters*, vol. 48, no. 6, pp. 1174–1179, 2006.

[13] M Kössl and IJ Russell, "Basilar membrane resonance in the cochlea of the mustached bat," *Proc Natl Acad Sci U S A*, vol. 92, pp. 276–9, 1995.

[14] Shizuko Hiryu, Koji Katsura, Liang-Kong Lin, Hiroshi Riquimaroux, and Yoshiaki Watanabe, "Radiation pattern of echolocation pulse in Taiwanese leaf-nosed bat, hipposideros terasensis," *Acoustical Science and Technology*, vol. 27, no. 2, pp. 108–110, 2006.

[15] JM Wotton, RL Jenison, and DJ Hartley, "The combination of echolocation emission and ear reception enhances directional spectral cues of the big brown bat, eptesicus fuscus," *J Acoust Soc Am*, vol. 101, pp. 1723–33, 1997.

[16] JM Wotton, T Haresign, and JA Simmons, "Spatially dependent acoustic cues generated by the external ear of the big brown bat, eptesicus fuscus," *J Acoust Soc Am*, vol. 98, pp. 1423–45, 1995.

[17] J. F. Fish, C. S. Johnson, and D. K. Ljungblad, "Sonar target discrimination by instrumented human divers," *Journal of the Acoustical Society of America*, vol. 59, no. 3, pp. 602–606, 1976.

[18] Clarence F. Ramstedt, "Active sonar image perception," U.S. Patent #3887896, April 1974.

[19] Florian M. König, "A new supra-aural dynamic headphone system for in-front localization and surround reproduction of sound," in *AES Convention 1997 Preprint*, 1997.

[20] A Kulkarni and HS Colburn, "Role of spectral detail in sound-source localization," *Nature*, vol. 396, pp. 747–9, 1998.

[21] SensComp, "Piezo transducer 40LR16/40LT16 datasheet," `http://www.senscomp.com/specs/40LT16%20%20spec.pdf`, April 2004.

[22] SA Kick and JA Simmons, "Automatic gain control in the bat's sonar receiver and the neuroethology of echolocation," *J Neurosci*, vol. 4, pp. 2725–37, 1984.